



(12) 发明专利申请

(10) 申请公布号 CN 120126162 A

(43) 申请公布日 2025. 06. 10

(21) 申请号 202510197844.8

(22) 申请日 2025.02.21

(71) 申请人 内蒙古师范大学

地址 010020 内蒙古自治区呼和浩特市赛
罕区昭乌达路81号

(72) 发明人 董杰 李艳玲 郭浩林 杨承
白胡日查 王志伟

(74) 专利代理机构 西安智科树知识产权代理事
务所(普通合伙) 61289
专利代理师 张晓

(51) Int.Cl.

G06V 30/412 (2022.01)

G06V 30/19 (2022.01)

G06V 10/82 (2022.01)

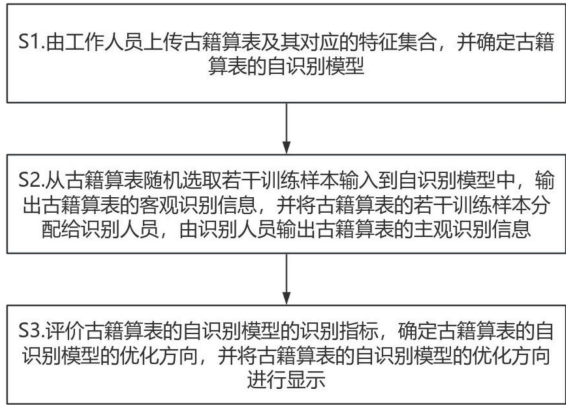
权利要求书3页 说明书6页 附图1页

(54) 发明名称

一种古籍算表识别评价方法、系统及电子设备

(57) 摘要

本发明公开了一种古籍算表识别评价方法、系统及电子设备,涉及古籍算表识别技术领域,本发明包括:S1.由工作人员上传古籍算表及其对应的特征集合,并确定古籍算表的自识别模型;S2.从古籍算表随机选取若干训练样本输入到自识别模型中,输出古籍算表的客观识别信息,并将古籍算表的若干训练样本分配给识别人员,由识别人员输出古籍算表的主观识别信息;S3.评价古籍算表的自识别模型的识别指标,确定古籍算表的自识别模型的优化方向,并将古籍算表的自识别模型的优化方向进行显示,本发明避免造成自识别模型的资源浪费,也提高古籍算表的识别效率,本发明确保了古籍算表在文字识别、数值识别等的精度。



1. 一种古籍算表识别评价方法,其特征在于,包括:

S1. 由工作人员上传古籍算表及其对应的特征集合,并确定古籍算表的自识别模型;

所述自识别模型包括图像处理方法和识别方法的搭配;

S2. 从古籍算表随机选取若干训练样本输入到自识别模型中,输出古籍算表的客观识别信息,并将古籍算表的若干训练样本分配给识别人员,由识别人员输出古籍算表的主观识别信息;

S3. 评价古籍算表的自识别模型的识别指标,确定古籍算表的自识别模型的优化方向,并将古籍算表的自识别模型的优化方向进行显示。

2. 根据权利要求1所述的一种古籍算表识别评价方法,其特征在于,所述确定古籍算表的自识别模型,其具体确定过程为:

S100、通过图像识别技术识别古籍算表的缺陷参数;

所述缺陷参数为缺陷总体表征参量;

S101、将古籍算表的缺陷参数补充到古籍算表的特征集合中,得到古籍算表的更新特征集合;

S102、将古籍算表的更新特征集合与web数据仓中存储的古籍算表的难易程度划分表进行对比,经匹配得到古籍算表的难易程度;

S103、将古籍算表的难易程度与web数据仓中存储的古籍算表的各难易程度对应的自识别模型进行对比,经匹配得到古籍算表的自识别模型。

3. 根据权利要求1所述的一种古籍算表识别评价方法,其特征在于,所述客观识别信息包括若干训练样本的文本数据集、数值数据集、符号数据集和算表结构数据集;

所述主观识别信息包括若干训练样本的文本数据集、数值数据集、符号数据集和算表结构数据集。

4. 根据权利要求3所述的一种古籍算表识别评价方法,其特征在于,所述评价古籍算表的自识别模型的各级识别指标,其具体实现步骤为:

S300、基于古籍算表的客观识别信息和主观识别信息,以某训练样本为指定样本,确定古籍算表的指定样本的第一识别指标 α_1 、第二识别指标 α_2 、第三识别指标 α_3 和第四识别指标 α_4 ;

S301、以此类推,得到古籍算表的各训练样本的第一识别指标、第二识别指标、第三识别指标和第四识别指标,并将其进行均值处理,得到古籍算表的第一识别指标均值、第二识别指标均值、第三识别指标均值和第四识别指标均值,作为古籍算表的自识别模型的一级识别指标 β_1 、二级识别指标 β_2 、三级识别指标 β_3 、四级识别指标 β_4 。

5. 根据权利要求4所述的一种古籍算表识别评价方法,其特征在于,所述确定古籍算表的指定样本第一识别指标 α_1 ,其具体确定方法为:

从古籍算表的指定样本的客观识别信息中提取文本数据集中的表头关键字集合和各单元格的关键字集合,并从古籍算表的指定样本的主观识别信息中提取文本数据集中的表头关键字集合和各单元格的关键字集合;

通过集合相似度算法,计算得到古籍算表的指定样本的客观识别信息中表头关键字集合与主观识别信息中表头关键字集合的相似度,作为古籍算表的指定样本的表头识别准确度;

通过集合相似度算法,计算得到古籍算表的指定样本的客观识别信息中各单元格的关键字集合与对应单元格的关键字集合的相似度,作为古籍算表的指定样本的各单元格的识别准确度,筛选古籍算表的指定样本的各异质单元格;

汇总古籍算表的指定样本的异质单元格的总数量 M' 和单元格的总数量 M' ,并将其与古籍算表的指定样本的表头识别准确度 ε' 导入到第一识别指标模型

$$\alpha_{-1} = \ln \left(1 + \varepsilon' * \frac{M'}{1 + M} \right) \text{ 中, 输出古籍算表的指定样本的第一识别指标。}$$

6. 根据权利要求4所述的一种古籍算表识别评价方法,其特征在于,所述确定古籍算表的指定样本的第三识别指标 α_{-3} ,其具体确定方法为:

从古籍算表的指定样本的客观识别信息中提取符号数据集中的运算符集合、表格分隔符号集合,并从古籍算表的指定样本的主观识别信息中提取符号数据集中的运算符集合、表格分隔符号集合;

通过集合相似度算法,计算古籍算表的指定样本的客观识别信息中运算符集合与主观识别信息中运算符集合的相似度,作为古籍算表的指定样本的运算符识别准确度,同理,计算得到古籍算表的指定样本的表格分隔符号识别准确度;

将古籍算表的运算符识别准确度 η 和表格分隔符号识别准确度 μ 导入到第三识别指标模型 $\alpha_{-3} = \begin{cases} 1, (\eta \geq \eta') \wedge (\mu \geq \mu') \\ 0, (\eta < \eta') \vee (\mu < \mu') \end{cases}$ 中,输出古籍算表的指定样本的第三识别指标,式

中 η' 、 μ' 分别表示为web数据仓库中存储的运算符识别准确度阈值、表格分隔符号识别准确度阈值, $\wedge$ 、 $\vee$ 分别为逻辑符号且、或。

7. 根据权利要求4所述的一种古籍算表识别评价方法,其特征在于,所述确定古籍算表的指定样本的第四识别指标 α_{-4} ,其具体确定方法为:

从古籍算表的指定样本的客观识别信息中提取算表结构数据集的行数、列数、各合并单元格和各拆分单元格,并从古籍算表的指定样本的主观识别信息中提取算表结构数据集的行数、列数、各合并单元格和各拆分单元格,分别记为古籍算表的参考行数、参考列数、各参考合并单元格和各参考拆分单元格;

比对分析得到古籍算表的指定样本的行数偏差值、列数偏差值、各偏差合并单元格和各偏差拆分单元格,汇总得到古籍算表的指定样本的偏差合并单元格的总数量和偏差拆分单元格的总数量;

汇总得到古籍算表的指定样本的参考合并单元格的总数量和参考拆分单元格的总数量,经数值处理得到古籍算表的指定样本的第四识别指标,所述第四识别指标包含0和1的数值。

8. 根据权利要求3所述的一种古籍算表识别评价方法,其特征在于,所述确定古籍算表的自识别模型的优化方向,其具体确定方法为:

基于古籍算表的第一识别指标 α_{-1} 、第二识别指标 α_{-2} 、第三识别指标 α_{-3} 和第四识别指标 α_{-4} ;

若 $(\beta_{-1} < \beta'_{-1})$,则将古籍算表的自识别模型的优化方向记为文本优化;

若 $(\beta_{-2} < \beta'_{-2})$,则将古籍算表的自识别模型的优化方向记为数值优化;

若 $(\beta_3=0)$,则将古籍算表的自识别模型的优化方向记为符号优化;

若 $(\beta_4=0)$,则将古籍算表的自识别模型的优化方向记为算表结构优化;

β'_1 、 β'_2 分别为web数据仓中存储的一级识别指标收敛值、二级识别指标收敛值;

汇总得到古籍算表的自识别模型的优化方向。

9.一种用于执行权利要求1-8任一项所述古籍算表识别评价方法的系统,其特征在于,包括:

古籍算表上传模块,用于由工作人员上传古籍算表及其对应的特征集合,并确定古籍算表的自识别模型;

所述自识别模型包括图像处理方法和识别方法的搭配;

古籍算表识别模块,用于从古籍算表随机选取若干训练样本输入到自识别模型中,输出古籍算表的客观识别信息,并将古籍算表的若干训练样本分配给识别人员,由识别人员输出古籍算表的主观识别信息;

古籍识别评价模块,用于评价古籍算表的自识别模型的识别指标,确定古籍算表的自识别模型的优化方向,并将古籍算表的自识别模型的优化方向进行显示。

10.一种电子设备,其特征在于,包括:处理器,存储器及通信总线;所述存储器上存储有可被所述处理器执行的计算机可读程序;所述通信总线实现处理器和存储器之间的连接通信;所述处理器执行所述计算机可读程序时实现如权利要求1-8任意一项所述的古籍算表识别评价方法。

一种古籍算表识别评价方法、系统及电子设备

技术领域

[0001] 本发明涉及古籍算表识别技术领域,具体涉及一种古籍算表识别评价方法、系统及电子设备。

背景技术

[0002] 古籍算表作为数学智慧的结晶,不仅记录了古代数学的具体计算方法,还反映了当时数学发展的水平、特点和趋势。通过对算表的研究,可以深入了解古代数学的起源、演变和成就,为现代数学史的研究提供实物证据和理论支撑。通过识别与评价,可以对这些珍贵的文化遗产进行系统的整理和保护,避免因时间、环境等因素导致的信息丢失或损坏,确保数学知识的传承不断档,因此,对古籍算表识别评价是极其有必要的。

[0003] 而对于目前的古籍算表识别,现有技术中还存在一些不足,具体可以体现在以下几个层面:现有技术中鲜少基于古籍算表及其对应的特征集合,确定古籍算表的自识别模型,从而难以保障古籍算表与自识别模型的初始适配性,容易造成自识别模型的资源浪费,也降低古籍算表的识别效率,同时鲜少有通过古籍算表的自识别模型对训练样本进行识别后评价,而对古籍算表的自识别模型的优化方向进行确认,难以确保古籍算表的自识别模型的识别精准度,难以捕捉古籍算表的复杂特征。

发明内容

[0004] 本发明的目的在于提供的一种古籍算表识别评价方法、系统及电子设备,解决了背景技术中存在的问题。

[0005] 为解决上述技术问题,本发明采用如下技术方案:本发明第一方面提供一种古籍算表识别评价方法,包括:S1.由工作人员上传古籍算表及其对应的特征集合,并确定古籍算表的自识别模型。

[0006] 所述自识别模型包括图像处理方法和识别方法的搭配。

[0007] S2.从古籍算表随机选取若干训练样本输入到自识别模型中,输出古籍算表的客观识别信息,并将古籍算表的若干训练样本分配给识别人员,由识别人员输出古籍算表的主观识别信息。

[0008] S3.评价古籍算表的自识别模型的识别指标,确定古籍算表的自识别模型的优化方向,并将古籍算表的自识别模型的优化方向进行显示。

[0009] 本发明第二方面提供一种用于执行本发明所述古籍算表识别评价方法的系统,包括:古籍算表上传模块,用于由工作人员上传古籍算表及其对应的特征集合,并确定古籍算表的自识别模型。

[0010] 所述自识别模型包括图像处理方法和识别方法的搭配。

[0011] 古籍算表识别模块,用于从古籍算表随机选取若干训练样本输入到自识别模型中,输出古籍算表的客观识别信息,并将古籍算表的若干训练样本分配给识别人员,由识别人员输出古籍算表的主观识别信息。

[0012] 古籍识别评价模块,用于评价古籍算表的自识别模型的识别指标,确定古籍算表的自识别模型的优化方向,并将古籍算表的自识别模型的优化方向进行显示。

[0013] 本发明第三方面提供一种电子设备,包括:处理器,存储器及通信总线。所述存储器上存储有可被所述处理器执行的计算机可读程序。所述通信总线实现处理器和存储器之间的连接通信。所述处理器执行所述计算机可读程序时实现如本发明所述的古籍算表识别评价方法。

[0014] 本发明的有益效果在于:(1)本发明基于古籍算表及其对应的特征集合,确定古籍算表的自识别模型,从而保障古籍算表与自识别模型的初始适配性,避免造成自识别模型的资源浪费,也提高古籍算表的识别效率,为后续古籍算表的自识别模型的优化方向的确认奠定基础。

[0015] (2)本发明通过古籍算表的自识别模型对训练样本进行识别后评价,而后对古籍算表的自识别模型的优化方向进行确认,在一定程度上提高了古籍算表的自识别模型的识别精准度,提高古籍算表复杂特征的捕捉率,确保古籍算表在文字识别、数值识别等的精度。

附图说明

[0016] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0017] 图1为本发明方法实施步骤流程示意图。

[0018] 图2为本发明系统结构连接示意图。

具体实施方式

[0019] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0020] 参照图1所示,本发明第一方面提供一种古籍算表识别评价方法,包括:S1.由工作人员上传古籍算表及其对应的特征集合,并确定古籍算表的自识别模型。

[0021] 需要说明的是,所述特征集合包括内容复杂特征值、表格结构复杂特征值、字体复杂特征值等,所述内容复杂特征值、表格结构复杂特征值和字体复杂特征值具体为0-1的数值。

[0022] 所述自识别模型包括图像处理方法和识别方法的搭配。

[0023] 在本发明的具体实施例中,所述确定古籍算表的自识别模型,其具体确定过程为:S100、通过图像识别技术识别古籍算表的缺陷参数。

[0024] 所述缺陷参数为缺陷总体表征参量。

[0025] 需要说明的是,所述缺陷总体表征参量具体为均一化处理后的缺陷总面积,例如缺陷总面积为10平方米,古籍算表的总面积为80平方米,则对于0、10、80进行均一化处理,

则均一化处理后的缺陷总面积为0.125。

[0026] S101、将古籍算表的缺陷参数补充到古籍算表的特征集合中,得到古籍算表的更新特征集合。

[0027] S102、将古籍算表的更新特征集合与web数据仓中存储的古籍算表的难易程度划分表进行对比,经匹配得到古籍算表的难易程度。

[0028] 需要说明的是,所述古籍算表的难易程度划分表为各难易程度对应的特征集合中各元素的需求区间,例如内容复杂特征值为0.1、表格结构复杂特征值为0.1、字体复杂特征值为0.1、缺陷总体表征参量为0.2,处于古籍算表的容易程度对应特征集合中各元素的需求区间内,则将古籍算表的难易程度记为容易。

[0029] S103、将古籍算表的难易程度与web数据仓中存储的古籍算表的各难易程度对应的自识别模型进行对比,经匹配得到古籍算表的自识别模型。

[0030] 需要说明的是,所述古籍算表的各难易程度对应的自识别模型,例如,古籍算表的难易程度为容易,则图像处理为去噪+二值化+倾斜校正,识别方法为:传统OCR+规则-based表格结构识别,若古籍算表的难易程度为困难,则图像处理为超分辨率+对比度增强+表格区域检测,识别方法为深度学习OCR(如CRNN或TrOCR)+深度学习表格结构识别(如TableNet),具体由工作人员上传。

[0031] 本发明基于古籍算表及其对应的特征集合,确定古籍算表的自识别模型,从而保障古籍算表与自识别模型的初始适配性,避免造成自识别模型的资源浪费,也提高古籍算表的识别效率,为后续古籍算表的自识别模型的优化方向的确认奠定基础。

[0032] S2.从古籍算表随机选取若干训练样本输入到自识别模型中,输出古籍算表的客观识别信息,并将古籍算表的若干训练样本分配给识别人员,由识别人员输出古籍算表的主观识别信息。

[0033] 在本发明的具体实施例中,所述客观识别信息包括若干训练样本的文本数据集、数值数据集、符号数据集和算表结构数据集。

[0034] 所述文本数据集包括表头关键字集合和各单元格的关键字集合,所述符号数据集包括运算符号集合、表格分隔符号集合,所述算表结构数据集包括行数、列数、各合并单元格和各拆分单元格。

[0035] 所述主观识别信息包括若干训练样本的文本数据集、数值数据集、符号数据集和算表结构数据集。

[0036] S3.评价古籍算表的自识别模型的识别指标,确定古籍算表的自识别模型的优化方向,并将古籍算表的自识别模型的优化方向进行显示。

[0037] 在本发明的具体实施例中,所述评价古籍算表的自识别模型的各级识别指标,其具体实现步骤为:S300、基于古籍算表的客观识别信息和主观识别信息,以某训练样本为指定样本,确定古籍算表的指定样本的第一识别指标 α_1 、第二识别指标 α_2 、第三识别指标 α_3 和第四识别指标 α_4 。

[0038] 需要说明的是,所述确定古籍算表的指定样本的第二识别指标,其具体确定方法为:通过集合相似度算法,计算古籍算表的客观识别信息中数值数据集与主观识别信息中数值数据集的相似度,作为古籍算表的第二识别指标。

[0039] 还需要说明的是,所述集合相似度算法,具体为 $(X,S)=\frac{|A\cap B|}{|A\cup B|}$,其中A和B分别指代两个集合。

[0040] S301、以此类推,得到古籍算表的各训练样本的第一识别指标、第二识别指标、第三识别指标和第四识别指标,并将其进行均值处理,得到古籍算表的第一识别指标均值、第二识别指标均值、第三识别指标均值和第四识别指标均值,作为古籍算表的自识别模型的一级识别指标 β_1 、二级识别指标 β_2 、三级识别指标 β_3 、四级识别指标 β_4 。

[0041] 在本发明的具体实施例中,所述确定古籍算表的指定样本第一识别指标 α_1 ,其具体确定方法为:从古籍算表的指定样本的客观识别信息中提取文本数据集中的表头关键字集合和各单元格的关键字集合,并从古籍算表的指定样本的主观识别信息中提取文本数据集中的表头关键字集合和各单元格的关键字集合。

[0042] 通过集合相似度算法,计算得到古籍算表的指定样本的客观识别信息中表头关键字集合与主观识别信息中表头关键字集合的相似度,作为古籍算表的指定样本的表头识别准确度。

[0043] 通过集合相似度算法,计算得到古籍算表的指定样本的客观识别信息中各单元格的关键字集合与对应单元格的关键字集合的相似度,作为古籍算表的指定样本的各单元格的识别准确度,筛选古籍算表的指定样本的各异质单元格。

[0044] 需要说明的是,所述识别准确度具体为0-1的数值。

[0045] 还需要说明的是,所述筛选古籍算表的指定样本的各异质单元格,其具体筛选方法为:将古籍算表的指定样本的各单元格的识别准确度与web数据仓中存储的单元格识别准确度阈值进行对比,若某单元格的识别准确度小于单元格识别准确度阈值,则将该单元格记为异质单元格,筛选得到古籍算表的指定样本的各异质单元格,所述单元格识别准确度阈值具体由古籍算表的工作人员进行设置,例如为了提高古籍算表的文字识别准确度,将单元格识别准确度阈值设置为0.95。

[0046] 汇总古籍算表的指定样本的异质单元格的数目M和单元格的总数量M',并将其与古籍算表的指定样本的表头识别准确度 ε' 导入到第一识别指标模型

$$\alpha_1 = \ln \left(1 + \varepsilon' * \frac{M'}{1+M} \right) \text{ 中, 输出古籍算表的指定样本的第一识别指标。}$$

[0047] 需要说明的是,第一识别指标模型采用 $\ln()$ 函数是为了将取值范围控制在0-1范围之内,便于后续分析。

[0048] 在本发明的具体实施例中,所述确定古籍算表的指定样本的第三识别指标 α_3 ,其具体确定方法为:从古籍算表的指定样本的客观识别信息中提取符号数据集中的运算符集合、表格分隔符号集合,并从古籍算表的指定样本的主观识别信息中提取符号数据集中的运算符集合、表格分隔符号集合。

[0049] 通过集合相似度算法,计算古籍算表的指定样本的客观识别信息中运算符集合与主观识别信息中运算符集合的相似度,作为古籍算表的指定样本的运算符识别准确度,同理,计算得到古籍算表的指定样本的表格分隔符号识别准确度。

[0050] 将古籍算表的运算符识别准确度 η 和表格分隔符号识别准确度 μ 导入到第三识

别指标模型 $\alpha_{-3} = \begin{cases} 1, (\eta \geq \eta') \wedge (\mu \geq \mu') \\ 0, (\eta < \eta') \vee (\mu < \mu') \end{cases}$ 中, 输出古籍算表的指定样本的第三识别指

标, 式中 η' 、 μ' 分别表示为 web 数据仓中存储的运算符号识别准确度阈值、表格分隔符号识别准确度阈值, $\wedge$ 、 $\vee$ 分别为逻辑符号且、或。

[0051] 需要说明的是, 所述运算符号识别准确度阈值、表格分隔符号识别准确度阈值具体由古籍算表的工作人员进行设置, 例如为了提高古籍算表的运算符号识别准确度、表格分隔符号识别准确度, 将运算符号识别准确度阈值、表格分隔符号识别准确度阈值均设置为 0.95。

[0052] 在本发明的具体实施例中, 所述确定古籍算表的指定样本的第四识别指标 α_{-4} , 其具体确定方法为: 从古籍算表的指定样本的客观识别信息中提取算表结构数据集的行数、列数、各合并单元格和各拆分单元格, 并从古籍算表的指定样本的主观识别信息中提取算表结构数据集的行数、列数、各合并单元格和各拆分单元格, 分别记为古籍算表的参考行数、参考列数、各参考合并单元格和各参考拆分单元格。

[0053] 比对分析得到古籍算表的指定样本的行数偏差值、列数偏差值、各偏差合并单元格和各偏差拆分单元格, 汇总得到古籍算表的指定样本的偏差合并单元格的数量和偏差拆分单元格的数量。

[0054] 需要说明的是, 所述比对分析得到古籍算表的指定样本的行数偏差值、列数偏差值、各行与各列的逻辑反馈值、各偏差合并单元格和各偏差拆分单元格, 其具体比对过程为: 将古籍算表的指定样本的行数、列数分别与参考行数、参考列数相减, 得到古籍算表的指定样本的行数偏差值、列数偏差值, 将古籍算表的各合并单元格与各参考合并单元格进行对比, 若某合并单元格与各参考合并单元格均比对失败, 则将该合并单元格记为偏差合并单元格, 得到各偏差合并单元格, 同理, 得到各偏差拆分单元格。

[0055] 汇总得到古籍算表的指定样本的参考合并单元格的数量和参考拆分单元格的数量, 经数值处理得到古籍算表的指定样本的第四识别指标。

[0056] 需要说明的是, 所述处理得到古籍算表的指定样本的第四识别指标, 其具体处理方法为: 将古籍算表的指定样本的偏差合并单元格的数量除以参考合并单元格的数量、偏差拆分单元格的数量除以参考拆分单元格的数量, 分别得到古籍算表的指定样本的偏差合并单元格比例、偏差拆分单元格比例。

[0057] 将古籍算表的指定样本的行数偏差值、列数偏差值、偏差合并单元格比例、偏差拆分单元格比例分别与 web 数据仓中存储的允许行数偏差值、允许列数偏差值、偏差合并单元格比例阈值、偏差拆分单元格比例阈值进行对比, 若行数偏差值小于允许行数偏差值、列数偏差值小于允许列数偏差值、偏差合并单元格比例小于偏差合并单元格比例阈值且偏差拆分单元格比例小于偏差拆分单元格比例阈值, 则将古籍算表的指定样本的第四识别指标记为 1, 反之, 则将其记为 0。

[0058] 在本发明的具体实施例中, 所述确定古籍算表的自识别模型的优化方向, 其具体确定方法为: 基于古籍算表的第一识别指标 α_{-1} 、第二识别指标 α_{-2} 、第三识别指标 α_{-3} 和第四识别指标 α_{-4} 。

[0059] 若 $(\beta_{-1} < \beta'_{-1})$, 则将古籍算表的自识别模型的优化方向记为文本优化。

- [0060] 若 $(\beta_{-2} < \beta'_{-2})$, 则将古籍算表的自识别模型的优化方向记为数值优化。
- [0061] 若 $(\beta_{-3} = 0)$, 则将古籍算表的自识别模型的优化方向记为符号优化。
- [0062] 若 $(\beta_{-4} = 0)$, 则将古籍算表的自识别模型的优化方向记为算表结构优化。
- [0063] β'_{-1} 、 β'_{-2} 分别为web数据仓中存储的一级识别指标收敛值、二级识别指标收敛值。
- [0064] 需要说明的是, 所述第一识别指标收敛值、第二识别指标收敛值具体由古籍算表的工作人员进行设置, 例如为了保障古籍算表的自识别模型的文本识别效果和数值识别效果, 可以将第一识别指标收敛值、第二识别指标收敛值均设置为0.9。
- [0065] 汇总得到古籍算表的自识别模型的优化方向。
- [0066] 本发明通过古籍算表的自识别模型对训练样本进行识别后评价, 而后对古籍算表的自识别模型的优化方向进行确认, 在一定程度上提高了古籍算表的自识别模型的识别精准度, 提高古籍算表复杂特征的捕捉率, 确保古籍算表在文字识别、数值识别等的精度。
- [0067] 参照图2所示, 本发明第二方面提供一种用于执行本发明所述古籍算表识别评价方法的系统, 包括:
- [0068] 古籍算表上传模块, 用于由工作人员上传古籍算表及其对应的特征集合, 并确定古籍算表的自识别模型。
- [0069] 所述自识别模型包括图像处理方法和识别方法的搭配。
- [0070] 古籍算表识别模块, 用于从古籍算表随机选取若干训练样本输入到自识别模型中, 输出古籍算表的客观识别信息, 并将古籍算表的若干训练样本分配给识别人员, 由识别人员输出古籍算表的主观识别信息。
- [0071] 古籍识别评价模块, 用于评价古籍算表的自识别模型的识别指标, 确定古籍算表的自识别模型的优化方向, 并将古籍算表的自识别模型的优化方向进行显示。
- [0072] 需要说明的是, 本发明还包括web数据仓, 所述古籍算表上传模块与古籍算表识别模块连接, 古籍算表识别模块与古籍识别评价模块连接, web数据仓分别与古籍算表上传模块和古籍识别评价模块连接。
- [0073] 本发明第三方面提供一种电子设备, 包括: 处理器, 存储器及通信总线。所述存储器上存储有可被所述处理器执行的计算机可读程序。所述通信总线实现处理器和存储器之间的连接通信。所述处理器执行所述计算机可读程序时实现如本发明所述的古籍算表识别评价方法。
- [0074] 以上内容仅仅是对本发明的构思所作的举例和说明, 所属本技术领域的技术人员对所描述的具体实施例做各种各样的修改或补充或采用类似的方式替代, 只要不偏离发明的构思或者超越本发明所定义的范围, 均应属于本发明的保护范围。

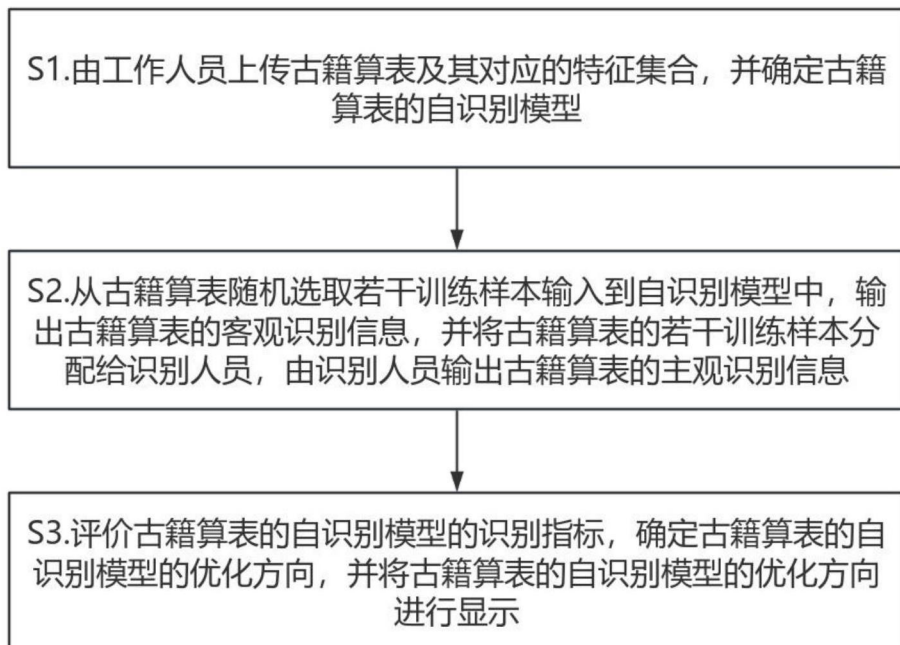


图1

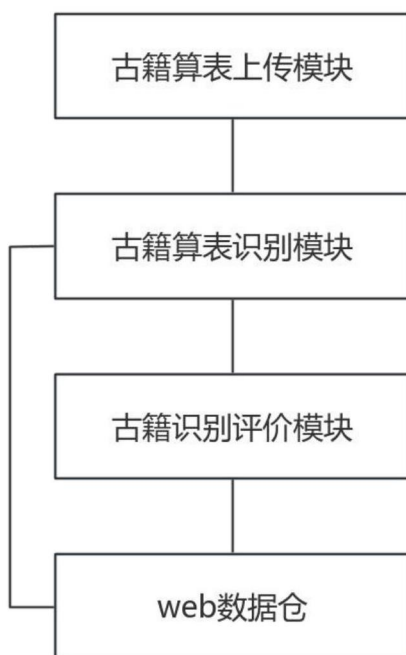


图2